How useful is the UMLS Metathesaurus in developing a controlled vocabulary for an automated problem list?

Thomas H. Payne, MD
David R. Martin, MD

Group Health Cooperative of Puget Sound
Seattle, Washington

ABSTRACT

We are developing a set of problem list phrases to be used in the automated problem list of a prototype clinical computing system. Because of the large number of terms in the Unified Medical Language System (UMLS™) and the links between them, we are experimenting with the use of the UMLS as the foundation for our problem list phrase set. We have found the UMLS to be very useful for this project, but that it lacks many phrases clinicians wish to include in the problem list. Internal linkages between phrases provided in the UMLS are not well suited to our needs. We plan to continue our use of the UMLS but to add problem list phrases and linkages between phrases to support browsing and decision support applications.

INTRODUCTION

A problem list is an important component of a clinical computing system. Clinicians use problem lists to summarize important patient characteristics and past medical problems. Problem list entries are also useful in decision support applications, and can be used to provide the reason for visit for billing purposes. For these reasons we are developing a problem list as part of a prototype clinical computing project at the Group Health Cooperative of Puget Sound.

Setting

The prototype clinical computing system is in use in one of Group Health Cooperative of Puget Sound's 26 clinics. It was developed internally over the last two years as part of an effort to upgrade clinical computing and to support population management in primary care. In addition to an automated problem list, the clinical computing system provides a current medication list, "advice modules," immunization history, listing of visits and procedures, and results of laboratory tests [1]. Physicians, nurse practitioners, nurses, medical assistants, and medical receptionists use workstations and paper encounter forms to view information on individual patients, or on the entire population (panel) of roughly 1500 patients served by the team.

In this paper we present our progress in using the UMLS as a starting point for developing the vocabulary for the problem list of our clinical computing system prototype. We review what we have learned about the usefulness and limitations of the UMLS for this purpose.

Desirable characteristics of a problem list vocabulary

The vocabulary used in the problem list of our clinical computing system should have the following characteristics:

Term acceptability. Clinician users must be satisfied with the terms in the vocabulary as substitutes for phrases they use in recording the problem list in the medical record.

Internal structure. Terms should have linkages to similar terms in the vocabulary. For example conditions affecting the same organ system, or which are broader or more specific examples of the same condition should be in some way linked together.

Map to other vocabularies. It should be possible to provide a representation of the problem list term in other vocabularies. Examples of the need for mapping are providing an ICD9-CM code for billing, or providing a MeSH term for literature searches.

Constraint There must be some constraint on the number of terms in the vocabulary, though the total number of phrases desired is not yet known.

In building our problem list vocabulary, we considered three options: using or adapting a single

0195-4210/92/$5.00

vocabulary such as ICD9-CM or SNOMED, developing an *ad hoc* vocabulary, or using the UMLS as our starting point.

Appeal of UMLS

We have previously studied the physician acceptance of International Classification of Diseases with Clinical Modifications, Ninth Revision (ICD9-CM) codes as substitutes for problem list phrases they record in the chart, and found that physicians are often dissatisfied with the ICD9-CM code substitute [2]. Since the UMLS Metathesaurus includes terms from many biomedical vocabularies including ICD9-CM, there is a greater likelihood that it contains a term which represents the term the clinician records in the chart in a form more acceptable to clinicians.

Within the UMLS there are linkages which might satisfy our need for internal structure for our problem list vocabulary. The concept hierarchy and the semantic network both provide linkages between the terms in the Metathesaurus. The context in which terms appear in the source vocabularies, and term co-occurrence with other terms in the biomedical literature indexing also provide potentially useful links [3].

An important practical point regarding development of a problem list is that the UMLS is easily available. Experimentation with its components is far easier because the most recent versions are available to registered users. New terms, links between terms, and edits are made regularly.

These characteristics of the UMLS led us to experiment with it to develop our problem list vocabulary.

METHODS

To begin construction of our problem list vocabulary we selected all terms in MRSO with source abbreviation = COS__, and stored the unique concept identifier (CUI), unique term identifier (LUI), and unique string identifier (SUI) for each row. We then obtained from MRCON the preferred form of the string corresponding to each CUI. The resulted in a set of 1472 phrases with the CUI, LUI, and SUI from the COSTAR set. We began with COSTAR terms because the COSTAR set was derived from a setting much like our own [4], and because it was of manageable size. We considered using SNOMED and ICD9 along with the COSTAR terms, but because there were 11,418 and 9,865 terms from SNOMED and ICD9 respectively, including them without restriction would have resulted in too large a phrase set for our initial experimentation in problem list vocabulary construction. We used Meta 1.2 in this study.

We expected that this set of COSTAR terms would

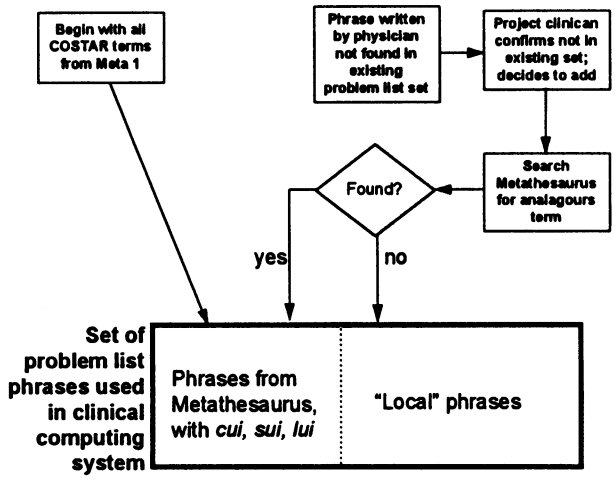

Figure 1. Method for adding phrase to problem list phrase set

not include many problem list phrases desired by clinicians, so we used the following method of entering additional terms to our set of problem list terms (see Figure 1):

Clinicians record problem list phrases on encounter forms used in the paper medical record. Data entry personnel attempt to find terms corresponding to each problem list phrase in the set of problem list phrases. If no corresponding phrase is found, a project clinician also searches the problem list set for a term. If she finds no term, she adds the new phrase to the problem list set. We then use the MetaCard™ application [5] to search for a corresponding term in the Metathesaurus. Though this process is continuing, we are developing three sets of phrases: those derived from the original COSTAR set from the Metathesaurus, phrases not in the original COSTAR set but which could be represented by a term in the Metathesaurus, and phrases which could not be represented by a term in the Metathesaurus, which we labeled "local" terms.

The result of this process is a set of phrases, some of which have a CUI corresponding to them. To support browsing by end users and for data entry, and for decision support applications, we have experimented with the use of the concept identifier as a link between similar terms. We have also explored the use of the UMLS semantic network for this purpose.

Related to the issue of providing links between similar problem list phrases is the need to provide modifiers to a base problem list phrase. Examples of modifiers we have found in use in problem list phrases from the medical record are terms expressing probability (R/O, probable), laterality (right, left), timing (H/P, chronic), location (thoracic, RLL), intensity (severe, mild), and number (multiple). The use of modifiers would potentially shorten the total number of phrases in the problem list vocabulary.

RESULTS

We have been adding to the problem list phrase set for two months. Additions to our problem list vocabulary are occurring daily, and have not yet begun to taper off as we expect to occur in the future.

We presently we have 1616 phrases in the problem list phrase set used in our clinical computing system prototype. The number of terms in each of these three categories is shown below:

Problem list phrase origin	Number
Phrases from Metathesaurus	
from COSTAR	1472
from other sources	60
"Local" phrases (not in Metathesaurus)	84
TOTAL PHRASES	1616

The 1472 COSTAR phrases included 33 duplicates (same phrase in COS89 and COS92). The remaining 1439 phrases represented 1361 unique UMLS concepts. Phrases with the same concept identifier were very similar, such as *aortic insufficiency*, and *regurgitation, aortic valve*.

Though at present terms derived from COSTAR represent the largest category, the other categories are growing rapidly. Examples of phrases in each of the three categories are:

Phrases from Metathesaurus:
from COSTAR
- Appendicitis
- Anticoagulation
- Elevated liver function tests
- Heart failure, congestive

from other sources
- Renal insufficiency
- Estrogen replacement
- Prostatic hypertrophy, benign
- Panic attacks

"Local" phrases (not in Metathesaurus)
- Health maintenance
- Myocardial infarction, anteroseptal
- Impingement syndrome
- Class III Pap smear
- Ureterolithotomy
- Elevated PSA
- Family history colon cancer
- Sedentary life style

An example of the need for links between terms is a decision support rule concerning influenza vaccine. One indication for influenza vaccine is a history of pulmonary disease. We have many terms for

pulmonary disease, very few of which share a concept identifier. Decision support rules which search for patients needing influenza vaccine must include a long list of codes for pulmonary disease. Since most of these pulmonary disorders fall into the "disease or syndrome" semantic type along with other diseases, the semantic network does not provide a useful means of grouping these pulmonary disorders together.

Our initial impression is that while both the semantic network and concept links are potentially useful, neither groups terms together in a fashion we need to allow browsing and for decision support. For example, we would like to be able to group problem list phrases by organ system, and all similar diagnoses, such as all problem list phrases for anemia. Our planned solution is to create *ad hoc* linkages between terms to supplement those provided in the UMLS, and to further explore the UMLS for other useful linkages.

Another need for our problem list vocabulary is to provide mapping to other vocabularies. The most pressing example of this need is to provide an ICD9-CM code for each problem list entry, so that if an entry is made at the time of an encounter, this entry can be used to generate an ICD9-CM code as the reason for visit for billing purposes. Although the UMLS provides the ability to translate many terms between vocabularies, for example from ICD9-CM to MeSH, it does not provide translation for all terms [6]. We generated an ICD9-CM code for each member of the problem list vocabulary using an automated coder (Autocoder, GMIS, Malvern, Pennsylvania). In cases where the automated coder could not provide a code, coding was done manually.

DISCUSSION

The purpose of the UMLS is to "facilitate the retrieval and integration of information from multiple machine-readable biomedical information sources" [7]. We have experimented in using it as a foundation for the problem list vocabulary for a clinical computing system. We have found the UMLS Metathesaurus to be a useful starting point for developing a set of phrases to be used in the problem list vocabulary for a clinical computing system problem list. The main advantages to using the UMLS are 1) the availability of a large set of terms from which to draw problem list terms; 2) links the UMLS provides to existing biomedical vocabularies including MeSH may be very useful in future clinical computing applications; 3) potential for comparisons between institutions provided by use of an internationally known system such as UMLS and 4) the advantage of ongoing maintenance of the UMLS which may provide more terms and links in the future.

Other groups have used the UMLS in developing links between patient records and bibliographic retrieval [8], between ICD9-CM diagnoses and procedures and MeSH [6].

We have found two main limitations to the UMLS for our purposes: A substantial number of terms clinicians wish to have in a problem list vocabulary are not found in the Metathesaurus, and the internal linkages provided in the UMLS are not well suited to our needs--the concept link is too fine and the semantic network links are too "coarse." The internal linkages provide only spotty resources for helping us to develop groupings of terms that will be useful in the clinical setting for browsing, automated decision support, and problem list management.

Analysis of UMLS coverage of concepts in hypertension notes [9], clinical radiology [10], and laboratory terminology [11] shows limitations of Metathesaurus terms in these areas as well. Although it is clear that we will have to add many terms specific to our clinical setting, there is much ongoing work in enhancing this aspect of the Metathesaurus.

Our initial approach to developing the problem list directory lacks adequate internal structure and other features of vocabularies elsewhere [12], but was relatively easy to develop with the aid of the Metathesaurus. We plan several extensions to our current set of problem list phrases. We will add additional terms, though we expect that the rate of growth of the problem list set will drop. We experimenting with several approaches to providing an internal structure to the problem list set, including adding links to indicate terms closely related, and which terms share a common organ system. We are also experimenting with an object oriented method of linking similar terms. We will

also explore further adapting links provided in the UMLS to our purposes, such as the semantic network.

We are using the Metathesaurus as a resource for terms and relationships that we find to be of value in our setting. We expect our set of terms to be small relative to the Metathesaurus, small enough to be handled easily at the individual clinical workstations. By starting with a set of terms from the Metathesaurus and by linking added terms as closely as possible to terms in the Metathesaurus, we hope to be able to maintain fairly close linkage with the Metathesaurus without having the higher overhead of maintaining a far larger set. As clinical terms become better represented in future versions of the Metathesaurus, we anticipate that more of our terms will find direct linkage [13].

Our goal is to develop a set of problem list phrases that clinicians find useful and familiar, which is well suited to decision support applications, and which provides links to other biomedical vocabularies used in the setting of our health maintenance organization and nationally in other organizations. The UMLS Metathesaurus has been very useful in our early work toward this goal.

REFERENCES

1. Payne TH, Savarino J. Computer "Advice" aids in managing care of a population. *HMO Practice* 1993 (in press).

2. Payne TH, Murphy GR, Salazar AS. How well does ICD9 represent phrases used in the medical record problem list? *Proceedings of the Sixteenth Annual SCAMC*, IEEE Computer Society Press, New York; November, 1992:654-657.

3. Nelson SJ, Tuttle MS, Cole WG, et al. From meaning to term: semantic locality in the UMLS Metathesaurus. *Proceedings of the Fifteenth Annual SCAMC*, IEEE Computer Society Press, New York; November, 1991:209-213.

4. Barnett GO. The application of computer-based medical-record systems in ambulatory practice. *NEJM* 1984; 310:1643-1650.

5. MetaCard™, a HyperCard interface to the NLM's UMLS Metathesaurus. National Library of Medicine. Documentation, 3rd Experimental Edition, UMLS™ Knowledge Sources, August, 1992

6. Cimino JJ, Johnson SB, Aguirre A. The Medline button. *Proceedings of the Sixteenth Annual SCAMC*, IEEE Computer Society Press, New York; November, 1992:81-85.

7. National Library of Medicine. Documentation, 3rd Experimental Edition, UMLS™ Knowledge Sources, August, 1992.

8. Powsner SM, Miller PL. From patient reports to bibliographic retrieval: A Meta-1 front-end. *Proceedings of the Sixteenth Annual SCAMC*, IEEE Computer Society Press, New York; November, 1991:526-530.

9. Campbell JR, Kallenberg GA, Sherrick RC. The clinical utility of META: an analysis for hypertension. *Proceedings of the Sixteenth Annual SCAMC* IEEE Computer Society Press, New York; November,1992:397-401.

10. Friedman C. The UMLS coverage of clinical radiology. *Proceedings of the Sixteenth Annual SCAMC*, IEEE Computer Society Press, New York; November, 1992:309-13.

11. Cimino JJ. Representation of clinical laboratory terminology in the Unified Medical Language System. *Proceedings of the Fifteenth Annual SCAMC*, IEEE Computer Society Press, New York; November, 1991: 199-203.

12. Cimino JJ, Hripcsac G, Johnson SB, Clayton PD. Designing an introspective, multipurpose, controlled medical vocabulary. *Proceedings of the Thirteenth Annual SCAMC*, IEEE Computer Society Press, New York; November, 1991:513-518..

13. Tuttle MS, Sherertz DD, Erlbaum MS, et al. Adding your terms and relationships to the UMLS Metathesaurus. *Proceedings of the Fifteenth Annual SCAMC*, IEEE Computer Society Press, New York; November, 1991:219-23